

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-242496

(43)Date of publication of application : 07.09.1999

(51)Int.Cl.

|         |       |
|---------|-------|
| G10L    | 3/00  |
| G10L    | 3/00  |
| G10L    | 9/00  |
| G11B    | 15/02 |
| G11B    | 27/10 |
| // G06F | 17/30 |

(21)Application number : 10-044151

(71)Applicant : KOBEL STEEL LTD

(22)Date of filing : 26.02.1998

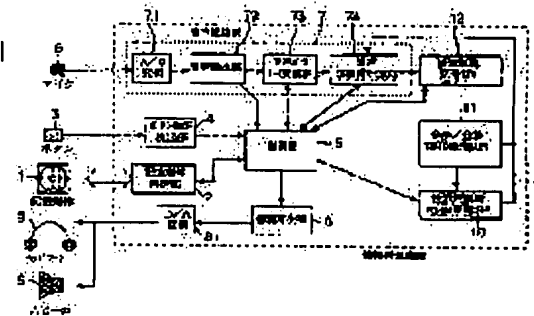
(72)Inventor : HIEKATA TAKAYUKI  
TAKAHASHI TETSUYA

**(54) INFORMATION REPRODUCING DEVICE**

**(57)Abstract:**

**PROBLEM TO BE SOLVED:** To easily and quickly retrieve information by constituting hidden Markov models for voice recognition while utilizing character information stored in a storage medium to eliminate a necessity in which a user registers voices for retrieval.

**SOLUTION:** When a storage medium 1 is inserted into a storage medium housing part 2, the insertion of a storing medium is detected by a control part 5 which is to be realized by calculating means of a CPU or the like and character information are read out from the medium 1. The read out character information are stored in a word HMM storage memory for retrieval 10 via the control part 5. Titles of musical compositions stored in the word HMM storage memory for retrieval 10 and HMMs stored in a phoneme/syllable HMM storage memory 11 (voice unit hidden Markov model means) every voice unit are combined by the control part 5 and HMM models corresponding to respective titles of musical compositions are plurally constituted every musical information. Values of similarity corresponding to respective musical information are stored in a retrieval history storage memory 12 by being arranged in the order of the high degrees of similarity.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

(19) 日本国特許庁 ( J P )

(12) 公開特許公報 ( A )

(11) 特許出願公開番号

特開平11-242496

(43) 公開日 平成11年 (1999) 9 月 7 日

(51) Int. Cl. <sup>6</sup>

識別記号

F I

G10L 3/00

551

G10L 3/00

551

G

535

551

Z

9/00

535

L

G11B 15/02

328

G11B 15/02

328

S

審査請求

未請求

請求項の数 6

〇 L

(全 7 頁)

最終頁に続く

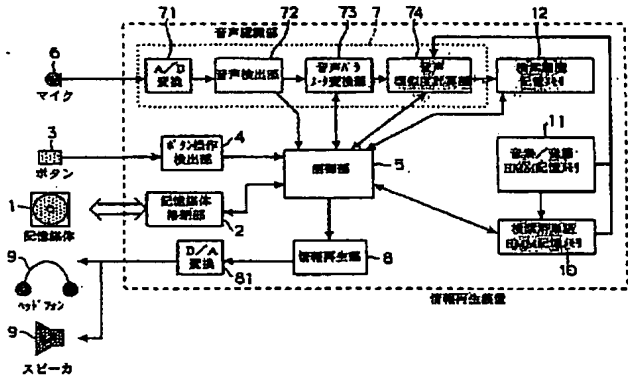
|           |                     |          |  |
|-----------|---------------------|----------|--|
| (21) 出願番号 | 特願平10-44151         | (71) 出願人 | 000001199<br>株式会社神戸製鋼所<br>兵庫県神戸市中央区脇浜町 1 丁目 3 番18号       |
| (22) 出願日  | 平成10年 (1998) 2 月26日 | (72) 発明者 | 稗方 孝之<br>兵庫県神戸市西区高塚台 1 丁目 5 番 5 号<br>株式会社神戸製鋼所神戸総合技術研究所内 |
|           |                     | (72) 発明者 | 高橋 哲也<br>兵庫県神戸市西区高塚台 1 丁目 5 番 5 号<br>株式会社神戸製鋼所神戸総合技術研究所内 |
|           |                     | (74) 代理人 | 弁理士 本庄 武男  |

(54) 【発明の名称】 情報再生装置

(57) 【要約】

【課題】 記憶媒体に複数記憶された楽曲等の情報を、音声によって選曲して再生を行う従来の再生装置では、予め楽曲名等を使用者の音声によって該記憶媒体に登録しておく必要があり、曲数が多い場合に登録作業が煩わしいものとなったり、書き込み不可の記憶媒体を用いて再生する装置には適用できない等の問題があった。

【解決手段】 本発明は、記憶媒体に記憶されている文字情報を利用して音声認識用のHMMを構成することにより、使用者が検索用の情報を登録する必要をなくして、所望の情報の検索を速やか且つ容易に行うことを図ったものである。



## 【特許請求の範囲】

【請求項 1】 記憶媒体に記憶された情報を入力された音声に基づいて選択し再生する情報再生装置であり、上記記憶媒体に記憶された情報を指定するための文字情報を入手する文字情報入手手段と、上記文字情報入手手段により入手された文字情報に対応する隠れマルコフモデルを構成する隠れマルコフモデル構成手段と、上記隠れマルコフモデル構成手段により構成された隠れマルコフモデルを用いて上記入力された音声に対応する上記文字情報を認識し、上記記憶媒体に記憶された情報を選択する音声認識情報選択手段とを具備してなる情報再生装置。

【請求項 2】 上記隠れマルコフモデル構成手段が、上記文字情報の文字単位に対応する所定の音声単位毎の音声単位隠れマルコフモデルが格納された音声単位隠れマルコフモデル記憶手段を具備し、上記音声単位隠れマルコフモデルを用いて上記文字情報に対応する上記隠れマルコフモデルを構成してなる請求項 1 に記載の情報再生装置。

【請求項 3】 上記文字情報入手手段が、上記情報に対応して予め上記記憶媒体に記憶されたインデックス情報から上記文字情報を入手してなる請求項 1 又は 2 に記載の情報再生装置。

【請求項 4】 上記文字情報入手手段が、上記情報が上記記憶媒体に記憶されている順番を基に上記文字情報を設定することにより、上記文字情報を入手してなる請求項 1 又は 2 に記載の情報再生装置。

【請求項 5】 上記隠れマルコフモデルと入力音声との類似度の大きい順に上記情報が再生されてなる請求項 1 ～ 4 のいずれか 1 項に記載の情報再生装置。

【請求項 6】 上記隠れマルコフモデルと入力音声との類似度が所定のしきい値以上の上記情報が、順次再生されてなる請求項 1 ～ 4 のいずれか 1 項に記載の情報再生装置。

## 【発明の詳細な説明】

## 【 0 0 0 1 】

【発明の属する技術分野】 本発明は、情報再生装置に係り、例えば音楽や映像等の情報を再生する情報再生装置に関するものである。

## 【 0 0 0 2 】

【従来の技術】 例えば MD、CD、半導体メモリ等の記憶媒体に記憶された音楽情報を再生する音楽情報再生装置では、該記憶媒体に複数記憶された音楽情報のいずれかを選択するのに、スキップボタンを用いて一曲づつ確認しながら検索を行ったり、多数設けられた数字ボタンにより曲番を入力して音楽情報の指定を行う場合が多い。ところで、近年の記憶媒体の大容量化、情報圧縮技術の進歩により、1つの記憶媒体に記憶することのできる楽曲数は増える傾向にあり、ボタンによる操作だけでは選曲に時間と手間を要するようになった。また、再生

装置自体も、ますます小型化する傾向にあり、特に携帯用の装置では、スペース的な制約により多くの操作ボタンを設けることが困難になり、選曲のための操作が複雑化した。このため、ボタンによる操作に加えて、音声認識により選曲操作を行う装置も開発されてきた。例えば特開昭 5 9 - 3 8 9 8 7 号公報には、音声入力により指定楽曲の自動演奏を行う技術が記載されている。この公報に記載の装置では、予め使用者により音声で入力された楽曲タイトルのパラメータが楽曲等の実データと共に上記記憶媒体に記憶され、検索時に上記楽曲タイトルが音声発生されれば、一致したインデックスに対応した実データが再生される。

## 【 0 0 0 3 】

【発明が解決しようとする課題】 しかしながら、上記従来の技術では、音声認識動作を行わせるために、使用者が予め音声登録を行う必要があった。特に全ての楽曲について音声認識検索しようとするれば、曲数だけの音声登録を行わなければならない、曲数が多くなるにつれて使用者の負担が増大した。また、記憶媒体毎に音声登録を行う必要があり、音声未登録の記憶媒体を用いる場合には、その記憶媒体に対して新たに音声登録を行わなければならない等の問題もあった。さらに、記憶媒体に検索用音声のパラメータを記憶させるため、CDやLD等の書き込みできない記憶媒体を用いた再生装置では使用することができないという問題もあった。また、検索用音声を登録する際と検索時の発声音声を入力する際との背景雑音の相違等のために、常に確実な音声認識を行うことは困難であり、認識率を向上させるために何回も音声登録を行う作業は使用者にとって煩わしいものであった。本発明は、このような従来の技術における課題を解決するために、情報再生装置を改良し、記憶媒体に記憶されている文字情報を利用して音声認識用のHMMを構成することにより、使用者が検索用音声を登録する必要をなくして、使用者の手を煩わせることなく、所望の情報の検索を速やか且つ容易に行う情報再生装置を提供することを目的とするものである。また、他の目的は、CD等の読み込み専用の記憶媒体を用いた場合でも、音声操作を行うことが可能な情報再生装置を提供することである。さらに、他の目的は、認識動作を何回もやり直す必要がない情報再生装置を提供することである。

## 【 0 0 0 4 】

【課題を解決するための手段】 上記目的を達成するために、請求項 1 に係る発明は、記憶媒体に記憶された情報を入力された音声に基づいて選択し再生する情報再生装置であり、上記記憶媒体に記憶された情報を指定するための文字情報を入手する文字情報入手手段と、上記文字情報入手手段により入手された文字情報に対応する隠れマルコフモデルを構成する隠れマルコフモデル構成手段と、上記隠れマルコフモデル構成手段により構成された隠れマルコフモデルを用いて上記入力された音声に対応

10

20

30

40

50

する上記文字情報を認識し、上記記憶媒体に記憶された情報を選択する音声認識情報選択手段とを具備してなる情報再生装置として構成されている。また、請求項 2 に係る発明は、上記請求項 1 に記載の情報再生装置において、上記隠れマルコフモデル構成手段が、上記文字情報の文字単位に対応する所定の音声単位毎の音声単位隠れマルコフモデルが格納された音声単位隠れマルコフモデル記憶手段を具備し、上記音声単位隠れマルコフモデルを用いて上記文字情報に対応する上記隠れマルコフモデルを構成してなることをその要旨とする。また、請求項 3 に係る発明は、上記請求項 1 又は 2 に記載の情報再生装置において、上記文字情報入手手段が、上記情報に対応して予め上記記憶媒体に記憶されたインデックス情報から上記文字情報を入手してなることをその要旨とする。また、請求項 4 に係る発明は、上記請求項 1 又は 2 に記載の情報再生装置において、上記文字情報入手手段が、上記情報が上記記憶媒体に記憶されている順番を基に上記文字情報を設定することにより、上記文字情報を入手してなることをその要旨とする。また、請求項 5 に係る発明は、上記請求項 1 ～ 4 のいずれか 1 項に記載の情報再生装置において、上記隠れマルコフモデルと入力音声との類似度の大きい順に上記情報が再生されてなることをその要旨とする。また、請求項 6 に係る発明は、上記請求項 1 ～ 4 のいずれか 1 項に記載の情報再生装置において、上記隠れマルコフモデルと入力音声との類似度が所定のしきい値以上の上記情報が、順次再生されてなることをその要旨とする。上記請求項 1 ～ 6 のいずれか 1 項に記載の情報再生装置によれば、上記記憶媒体に記憶されている上記文字情報を利用して音声認識用の HMM が構成されるため、使用者が、記憶媒体に検索用音声

【 0 0 0 5 】

【発明の実施の形態】以下、添付図面を参照して、本発

明の実施の形態につき説明し、本発明の理解に供する。この実施の形態では、本発明を例えば MD、CD 等の記憶媒体に記憶された音楽情報を再生する音楽情報再生装置に適用した例を示すが、これによって本発明の技術的範囲が限定されるものではない。本実施の形態に係る音楽情報再生装置は、上記記憶媒体に記憶された音声、楽器演奏等の音楽情報のうちのいずれかを、使用者の音声でもって指定して、聴取や検索のために再生するものであり、上記使用者の音声の認識には隠れマルコフモデル（以下、HMM と記す）が用いられる。また、上記音楽情報再生装置では、上記隠れマルコフモデルが、上記記憶媒体の各音楽情報に付加された楽曲順、楽曲名等の文字情報（インデックス情報）を基に構成される。まず、図 1 に上記音楽情報再生装置の概略構成を示す。上記音楽情報再生装置は、図 1 に示す如く、上記音楽情報が記憶された記憶媒体 1 が格納される記憶媒体格納部 2 と、再生、早送り等の動作に対応するボタン 3 の操作を検出するボタン操作検出部 4 と、上記記憶媒体格納部 2、ボタン操作検出部 4 等に接続され上記記憶媒体 1 に記憶された上記インデックス情報の検索、HMM の構成等の制御を行う制御部 5 と、上記記憶媒体 1 に記憶された音楽情報を指定するための音声が入力されるマイク等の音声入力部 6 と、上記制御部 5 により制御されて上記音声入力部 6 から入力された音声に対応する上記音楽情報を認識する音声認識部 7 と、上記音声認識部 7 により認識され上記制御部 5 により選択された音楽情報を再生する情報再生部 8 と、上記情報再生部 8 により再生される音楽情報を出力するヘッドフォン、スピーカ等の音声出力部 9 とを具備する。

【 0 0 0 6 】次に、上記音楽情報再生装置の詳細について、初期設定である検索用 HMM モデルの構成と、音声による音楽情報の検索動作という項目に分けて説明する。

（検索用 HMM モデルの構成）上記音楽情報再生装置では、音楽情報が記憶された例えば MD、CD 等の記憶媒体 1 が上記記憶媒体格納部 2 に挿入されると、CPU 等の演算手段により実現される制御部 5 により記憶媒体挿入が検知され、上記記憶媒体 1 から文字情報が読み出される。この文字情報には、例えば上記記憶媒体 1 に記憶されている楽曲名、アーティスト名、楽曲番号等のヘッダ、又はインデックス情報が利用される。読み出された文字情報は、制御部 5 を介して、検索用単語 HMM 記憶メモリ 10 に記憶される。この検索用単語 HMM 記憶メモリ 10 は、音素、音節等の音声単位毎に予め構成された HMM モデルが記憶される音素／音節 HMM 記憶メモリ 11 と共に、上記音声認識部 7 による音声認識の際に用いられる。例えば図 2 に示すように、記憶媒体 1 から「あかとんぼ」、「たきび」、「ちいさいあき」のような楽曲名が上記文字情報として読み出されると、音素／音節 HMM 記憶メモリ 11 に各音声単位毎に記憶された

HMMに対応する形で、音節HMMの例では、「A K A T O N B O」, 「T A K I B I」, 「T I I S A I A K I」というように、上記検索用単語HMM記憶メモリ10に楽曲名が記憶される。そして、上記音声認識部7による音声認識の際には、上記検索用単語HMMメモリ10に記憶された楽曲名と、上記音素／音節HMM記憶メモリ11（音声単位隠れマルコフモデル記憶手段）に各音声単位毎に記憶されたHMMが制御部5により組み合わせられて、楽曲名に対応するHMMモデルが各音楽情報毎に複数構成される。この制御部5及び音素／音節HMM記憶メモリ11による音声単位毎のHMMと検索用単語との組み合わせ動作が、本発明の隠れマルコフモデル構成手段に相当する。尚、上記記憶媒体1にヘッダ、又はインデックス情報が登録されていない場合には、例えば制御部5によって記憶媒体1に格納されている順番に音楽情報に番号を割り振り、この割り振りに用いた数字についてHMMを構成すればよい。即ち、上記記憶媒体1の一曲目に格納されている楽曲に対しては、「いち」という文字情報が作成され、これから「I I I」という形で、上記検索用単語HMMメモリ10に記憶される。作成される文字情報は、もちろん、「いち」という表現に限らず、「いちばんめ」、「いっきょくめ」、又は番号に対応づけられた単語等、順序を表す他の表現を用いてもよい。この制御部5による曲番からの文字情報設定動作、又は上記した記憶媒体1からの文字情報読み出し動作が、本発明の文字情報入手手段に相当する。このように、上記音楽情報再生装置では、上記記憶媒体1に記憶されている上記文字情報を利用して音声認識用のHMMが構成されるため、使用者が、記憶媒体1等に検索用音声新たに登録する必要がなく、何度も登録をやり直したり、記憶媒体1を交換する度に検索用音声を登録する等の煩雑な作業を行う必要がなくなる。このため、使用者の手を煩わせることなく、音声による検索に必要な音響モデルを作成することができる。また、HMMでは、確率モデルにより音声特徴量が表現されるため、話者の変動に対して柔軟な対応が可能である。また、検索用の音声を予め登録しておく必要がないので、音声登録用のメモリ等がない場合でも、CD等の読み込み専用の記憶媒体を用いることが可能である。

【0007】（音声による音楽情報の検索動作）図3に音声による音楽情報の検索動作を模式的に示す。尚、音声認識部7の構成については上記図1に示されている。上記音楽情報を指定するための音声は、マイク等の音声入力部6により上記音楽情報再生装置内に電気信号として取り込まれ、音声認識部7に出力される。入力された音声信号はA/D変換器71によってデジタル信号に変換される。上記A/D変換器71の後段には、音声検出器72が設けられている。この音声検出器72は、デジタル化された信号の中に発声された音声が含まれているか否かを検出する。上記音声検出器72における音

声の有無の検出は、例えば音量変動のパターンに基づいて行われる。この他、信号のゼロクロッシング、信号の周期性等、信号から抽出されるパターンに基づいた検出を行うものを用いることも、もちろん可能である。上記音声検出器72によって音声を検出されると、その後段にある音声パラメータ変換部73によって、上記音声について5～50mS程度のフレーム単位で音響分析が行われ、音声認識に有効な周波数パラメータが抽出される。このパラメータとしては、例えばLPCケプストラムやメル・ケプストラム等が用いられる。そして、音声類似度計算部74において、HMM毎に入力音声との類似度（尤度）の計算が行われる。このHMMが、上記検索用単語HMM記憶メモリ10に記憶された例えば楽曲名と、音素／音節HMM記憶メモリ11に記憶された各音声単位毎のHMMとを組み合わせることにより複数作成される、楽曲名のHMMである。各音楽情報に対応する類似度の値は、類似度の高いものから順序づけられて検索履歴記憶メモリ12に格納される。そして、例えば上記類似度が最も高い音楽情報が、上記制御部5により選択されて上記記憶媒体1から読み出され、上記情報再生部8によって再生され、D/A変換器81を介してヘッドフォン、スピーカ等の音声出力部9から出力される。尚、この音声認識部7による文字情報認識動作と制御部5による音楽情報選択動作が、音声認識情報選択手段に相当する。また、上記類似度が最も高い音楽情報だけを再生するのではなく、例えば上記類似度が高い順に上記音楽情報を順次再生し、再生、早送り、巻戻し、スキップ等に対応するボタンを用いて、所望の音楽情報を最終的に使用者が選択するようにしてもよい。この場合、環境条件等が悪く音声認識部で誤認識が行われた場合でも、次々に類似度が高い音楽情報が選択され、音声による再生動作をより安定にすることができる。このように、上記音楽情報再生装置によれば、上記記憶媒体に記憶されている上記文字情報を利用して音声認識用のHMMが構成されるため、各音楽情報毎に検索用音声を登録する必要もなく、さらに記憶媒体を交換しても速やかにHMMが構成される等、使用者の手を煩わせることなく、音声による再生音楽情報の検索が容易に行うことができる。また、複雑化する音楽情報検索のために再生、早送り等の基本的なボタンに加えて、新たに検索用のボタンを設ける必要がなくなるため、多くのボタンを設けることが困難な小型の情報再生装置でも良好な検索環境を提供することができる。さらに、類似度の高い順に音楽情報を順次再生すれば、環境条件等が悪く音声認識部における誤認識が行われるような場合でも、音声の認識動作を何回もやり直す必要がなく、音声による再生動作の円滑化を図ることができる。

【0008】

【実施例】上記実施の形態では、本発明を例えばMD、CD等の記憶媒体1に記憶された音楽情報を再生する音

10

20

30

40

50

楽情報再生装置に適用したが、これに限らず、LD、DVD、半導体メモリ等他の記憶媒体に記憶された音楽情報を再生する装置に適用してもよいし、音楽情報だけでなく、静止画像、ビデオ、文字情報等の他の情報を再生する装置に適用してもよい。このような情報再生装置も本発明における情報再生装置の一例である。また、上記実施の形態では、楽曲名や音楽情報が記憶されている順番に従って文字情報を入手しHMMを構成したが、この他、曲名、アーティスト名等の他の文字情報からHMMを構成するようにしてもよい。この場合、例えば所望のアーティストの音楽情報が同じ記憶媒体1に複数記憶されていれば、アーティスト名に基づくHMMにより複数の音楽情報が選択されることになるが、ボタン操作によって実際に再生する音楽情報を選択することも可能であるし、これらを連続して再生するようにしてもよい。このような情報再生装置も本発明における情報再生装置の一例である。また、上記実施の形態では、最も類似度の高い音楽情報のみを、又は類似度の高い順番に音楽情報を順次再生するようにしたが、これに限らず、例えば類似度が所定のしきい値以上の音楽情報を、検索で見つかった順番に再生するようにしてもよい。さらに、これらを組み合わせて、使用者が必要に応じて切り替えるように構成してもよい。また、上記実施の形態では、全ての情報の類似度の値を検索履歴記憶メモリに記憶するようにしたが、常に最も類似度の高い情報だけを記憶するようにしてもよい。このような情報再生装置も本発明における情報再生装置の一例である。また、上記実施の形態では、検索用のHMMを構成するために音素、音節毎に構成されたHMMを用いたが、音声単位としては、音声片（母音＋子音＋母音）等任意の単語列を形成できる他のものを用いてもよい。また、HMM自体にも様々な方式があるが、通常のHMMの他に、例えばATRにおいて提案されたHMnet等の方式を用いてもよい。このような情報再生装置も本発明における情報再生装置の一例である。尚、音声認識の方法としてHMMの代わりに、DPマッチングによる方法を用いることも原理的には可能であるが、DPマッチングでは音声特徴量が直接的に比較されるため、不特定話者の変動を吸収し得る柔

軟性がなく、装置を購入した使用者が数十～数百の音節・音素等を発声して登録する必要がある、現実的ではない。

【0009】

【発明の効果】上記のように上記請求項1～6のいずれか1項に記載の情報再生装置によれば、上記記憶媒体に記憶されている上記文字情報を利用して音声認識用のHMMが構成されるため、使用者が、記憶媒体に検索用音声新たに登録する必要がなく、何度も登録をやり直したり、記憶媒体を交換する度に検索用の情報を登録する等の煩雑な作業を行う必要がなくなる。このため、使用者の手を煩わせることなく、所望の情報の検索を速やか且つ容易に行うことが可能となる。その結果、再生、早送り等の基本的なボタンに加えて、新たに検索用のボタンを設ける必要がなくなるため、多くのボタンを設けることが困難な小型の装置に対しても利便性の高い検索環境を提供することができる。また、上記情報が上記記憶媒体に記憶されている順番を基に上記文字情報を設定することにより、CD等の読み込み専用の記憶媒体を用いることも可能となる。さらに、類似度の高い順に、又は所定のしきい値以上の情報を順次再生すれば、環境条件等が悪く誤認識が行われるような場合でも、認識動作を何回もやり直す必要がなく、再生動作の円滑化を図ることができる。

【図面の簡単な説明】

【図1】 本発明の一実施の形態に係る情報再生装置の概略構成を示す図。

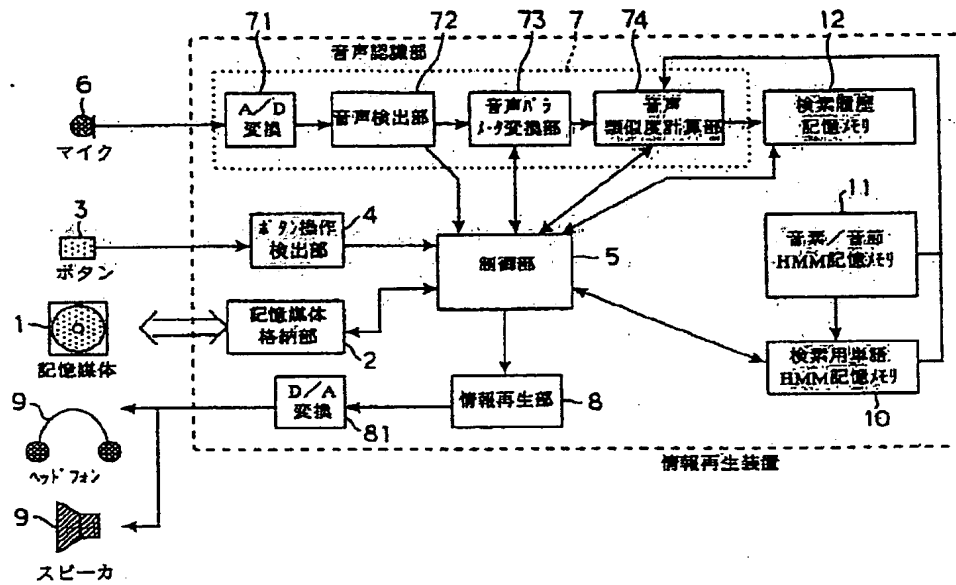
【図2】 上記情報再生装置における初期設定を模式的に示す図。

【図3】 上記情報再生装置における検索動作を模式的に示す図。

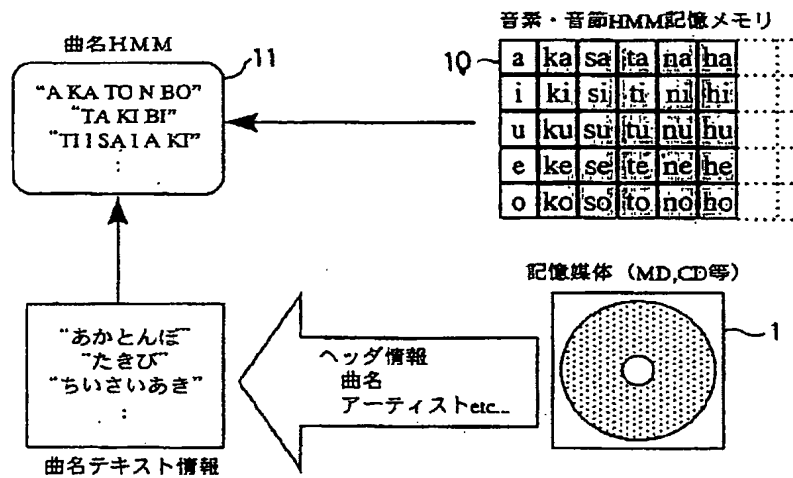
【符号の説明】

- 1…記憶媒体
- 5…制御部
- 7…音声認識部
- 1.1…音素／音節HMM記憶メモリ（音声単位隠れマルコフモデル記憶手段）

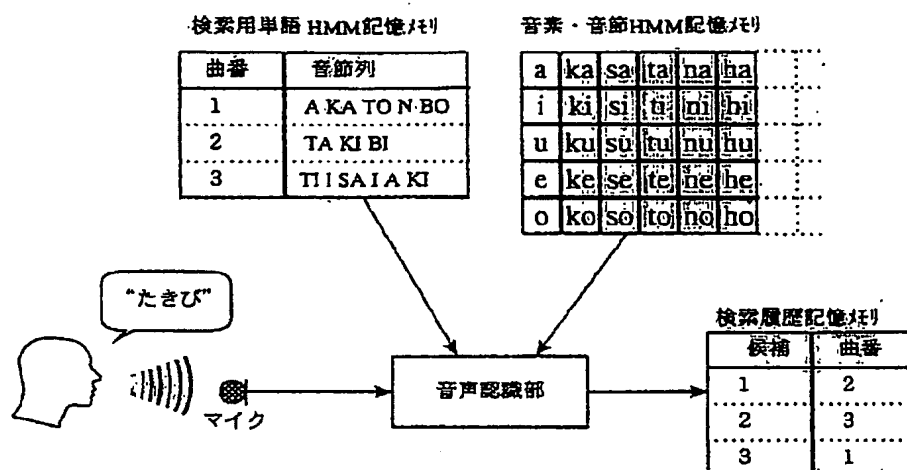
【図1】



【図2】



【図 3】



フロントページの続き

(51) Int. Cl. <sup>6</sup>

識別記号

F I

G 1 1 B 27/10

G 1 1 B 27/10

A

// G 0 6 F 17/30

G 0 6 F 15/40

3 7 0 E